

Development of Procedures to Assess Problem-Solving Competence in Computing Engineering

Jorge Pérez, Carmen Vizcarro, Javier García, Aurelio Bermúdez, and Ruth Cobos

Abstract—In the context of higher education, a competence may be understood as the combination of skills, knowledge, attitudes, values, and abilities that underpin effective and/or superior performance in a professional area. The aim of the work reported here was to design a set of procedures to assess a transferable competence, i.e., problem solving, that is basic for learning, in both academic and professional life, and crucial for engineering. The study involved a total of 71 students enrolled at three universities at two different stages of their studies. The development phases of the assessment device included an analysis of the competence and its facets, the design of the assessment task, the development of criteria to rate student performance, and the analysis of the basic psychometric properties for assessment methods in the area of education. The conclusion was drawn that the training process and the elaboration of scoring criteria are costly but necessary if objectivity in the interpretation of results is to be guaranteed. The main achievement of this project was the development of a procedure that measures learning outcomes and, more specifically, problem solving.

Index Terms—Engineering education, engineering education assessment, learning outcome (LO) assessment, problem solving, transferable competences.

I. INTRODUCTION

WHEN trying to assess student performance, teachers are interested in evaluating not only knowledge, as in traditional education, but also what the student is able to do with this knowledge (and how). Thus, the measurement of competences or learning outcomes (LOs) includes, but is not limited to, the assessment of knowledge [1]. This is normally evaluated through having the student carry out complex discipline-related tasks that require knowledge; such evaluations are often complemented with students' reflections.

Manuscript received October 28, 2015; revised April 8, 2016 and May 16, 2016; accepted June 5, 2016. This work was supported in part by the MINECO and European Commission (FEDER funds) under Project TIN2015-66972-C5-2-R; by the Spanish National Plan of R+D under Project TIN2014-52129-R; and by the Autonomous Community of Madrid under e-Madrid Project S2013/ICE-2715.

J. Pérez and J. García are with the Department of Computer Systems, Technical University of Madrid, 28031 Madrid, Spain (e-mail: jeperez@etsisi.upm.es; jgarcia@etsisi.upm.es).

C. Vizcarro, deceased, was with the Universidad Autónoma de Madrid, 28049 Madrid, Spain.

A. Bermúdez is with the Department of Computer Systems, University of Castilla-La Mancha, 02071 Albacete, Spain (e-mail: aurelio.bermudez@uclm.es).

R. Cobos is with the Universidad Autónoma de Madrid, 28049 Madrid, Spain (e-mail: Ruth.Cobos@uam.es).

Digital Object Identifier 10.1109/TE.2016.2582736

The assessment of LOs is a key concept in the European Credit Transfer and Accumulation System (ECTS) since credits are awarded when evaluation shows that the competences aimed at have been achieved. Most countries that adhere to the Bologna Process have adopted ECTS by law for their higher education systems [2]. In Spain, a Royal Decree [3] established five basic transferable competences that students of any university Bachelor's degree program must develop by the end of their studies. These include, in the student's specific fields of study, understanding basic and gradually more advanced scientific texts; problem solving; looking for, selecting, and using information to solve problems or make decisions; communication skills; and, finally, the capacity to learn independently. Competence-based education (CBE) and LO assessment are of interest not only to European or Spanish higher education professionals; there is general interest at an international level, as shown by projects such as Assessment of Learning Outcomes in Higher Education (AHELO) funded by the Organisation for Economic Co-operation and Development [4].

This paper is focused on the problem-solving competence. According to the ABET, the ability to identify, formulate, and solve engineering problems is among the 11 student outcomes that an engineering degree must ensure [5]. Australian studies have also identified problem solving as an important skill for engineers [6]. There is widespread agreement on its importance in the employability of engineers [7].

Several approaches have been taken to the theoretical principles underlying the field of CBE; these can be summarized as two extremes of a continuum. The origins of the work on CBE span from the mid-1950s to the 1970s, and the work of authors such as Bloom [8], [9] and, in the field of evaluation, Popham [10]. These authors emphasized the need for clear identification of the aims of a training program in observable terms. They favored operative definitions such as "*is able to list, calculate. . . and so on.*" Some of the ideas they introduced, including Bloom's taxonomy of learning objectives, are still influential today. However, this approach has been criticized on several grounds. The first of these is due to its atomistic nature. If all relevant educational objectives must be precisely operationalized, this ends up with long lists that then need to be further specified and described so that the desired fine detail can be reached. In practice, this means endless lists of very atomized descriptions that are difficult to understand by their users; both teachers and students are of little use in day-to-day educational work. Moreover and very significantly given the supposed objectivity of this approach, descriptions often

contain terms such as “adequately” or “timely” to describe the performance of behavior “X.” Of course, the question is what exactly do these terms mean [11], [12]. This approach may be adequate for descriptions of low-level skills, but less so for complex competences with high cognitive content, typical in the realm of higher education [13].

In contrast to this atomistic approach, a more holistic approach has been developed recently that seems better suited to tertiary education. Authors such as Bowden and Marton [11] and Stephenson [12] focus on rather complex developments that allow students to understand, think, and act in ways appropriate to the changing demands of the situations they face. While authors in these traditions agree on the value of describing the objectives from the very outset of an educational program, they consider that the aim of higher education today is not only to develop certain clear-cut competences, but rather to build the capacity to learn throughout life and to be able to cope with uncertain and unpredictable situations in unknown contexts. The educational value of this approach lies in its emphasis on the process through which complex competences are developed. From the point of view of assessment [14], however, it is understood that an effort must be made to achieve an appropriate level of objectivity.

CBE is a valid educational model in higher education because it promotes individualized learning while accommodating multiple learning styles and levels of engagement, as well as providing flexibility and affordability for students, among other benefits [15]. Edward and Knight [16] state that competences can be understood in operative terms and are directly connected with concrete practices. In the present work, this lies between the extremes of the atomistic and holistic traditions; the authors have identified activities with various degrees of complexity that are both meaningful from a disciplinary point of view and appropriate for the developmental level of the students concerned. An effort was also made to give precise descriptions of the achievable performance at different ability levels to provide adequate standards of objectivity in the assessment.

A. Problem-Solving Assessment Background

Problem solving can be described as the “*behavior and thought process directed toward the performance of some intellectually demanding task*” [17]. Many would consider this as one of the main goals of education. It is not surprising that human problem solving is a topic that has attracted much attention both in cognitive psychology and education. However, there is some controversy regarding whether problem solving can be taught as a general skill or needs to be imparted in specific knowledge domains. There is no doubt that domain knowledge is crucial to good performance, as the work of Polya [18] shows in mathematics, but there are also grounds for teaching problem solving as a general ability to be applied in specific knowledge contents [19], [20]. The work described here focused on a mathematical problem taught in a mathematics course. However, no attempt was made to teach this skill in a uniform way across all participating classes, with the tutors involved deciding for themselves which teaching approach to take.

TABLE I
SAMPLE DESCRIPTION BY UNIVERSITY, GRADE,
AGE, AND GENDER OF THE PARTICIPANTS

UNIVERSITY (sample %)	Gender (N - (%))		Age (mean)	Year (%)	
	Male	Female		1	3
UAM (18.46)	9 - (75)	3 - (25)	21.5	41.7	58.3
UCLM (36.92)	21 - (87.5)	3 - (12.5)	20.21	66.7	33.3
UPM (44.61)	26 - (89.65)	3 - (10.35)	18.31	82.9	17.1
TOTAL	N=56	N=9	20	63.76	36.23

In [21], the authors used a test based on the work of Tasa *et al.* [22] to evaluate the group work competence. In [23], the Problem Solving Inventory test [24] was used to measure the problem-solving competence. The drawback of approximations based on tests is that these techniques assess the student perception (subjective) about its own performance level in this competence, instead of its real performance level (objective) of the competence itself. Other work proposes tasks that focus on the different aspects comprising a given competence. The way in which the subject executes those tasks allows his or her expertise in the competence to be evaluated. For example, in [25], a task was developed to assess the competence related to understanding scientific texts.

B. Hypothesis

The working hypothesis of this paper is that it is possible to develop tasks that correctly represent the ability to solve problems. This is done by having students solve open-ended situations by using their background knowledge. This should be complemented by the development of assessment criteria that would provide enough objectivity when correcting and eventually grading students’ work.

II. METHOD

This section describes the development of a procedure to assess the basic transferable competence of problem solving. Students were set a task (see the Appendix) that consisted of a problem statement and a set of questions that measured students’ problem-solving ability. In addition, described is how basic objectivity and valid data for the assessment procedure were assured. Finally, some results are given on how the two academic levels participating in the study compare. Other comparisons of interest are the internal consistency of the test and closed versus open questions.

A. Participants

The study involved a total of 71 computer engineering students aged between 18 and 27. Table I gives the distribution of these students across the universities participating in the study, namely, the Universidad Autónoma de Madrid (UAM), the Universidad Politécnica de Madrid (UPM), and the Universidad de Castilla-La Mancha (UCLM). For six of the students, neither gender nor age is known.

In designing the task, an aim was that it should be useful for measuring the progress in the performance of the competence (discriminant validity). The initial plan was to sample first- and fourth-year students, but in Spanish universities, the fourth- and

final-year studies are entirely composed of elective courses, which could bias the sample. To avoid this, the sample was taken from first- and third-year students.

B. Procedure

In line with the standard procedures described in the literature [26], the design of these assessment procedures included: 1) a detailed analysis of the facets of each competence; 2) the design of assessment task covered these competences and their specific facets and have various levels of difficulty to accommodate the development of the student over the two academic years covered by the study; 3) the development of assessment criteria with an acceptable inter-rater reliability when grading students' work; and 4) determination of the basic psychometric properties that any measurement device should show, such as inter-rater agreement, internal consistency, and validity (content and discriminant validity), following the standards for educational and psychological testing [27].

1) *Curriculum Analysis and Student Recruitment*: The curriculum of the study program at the three universities was analyzed to find common content that could be used in the first and third years. The tests were taken by student volunteers enrolled in those academic years.

2) *Task Design*: To develop a task appropriate to comprehensively measure how LOs are associated with transferable competences, the competences were first analyzed in terms of their component parts, i.e., the aspects present in a problem-solving task analysis. The test questions were then mapped to this scheme: 1) identifying the problem; 2) creating a strategy to solve the problem; 3) finding additional information if necessary; 4) applying knowledge needed for problem solving; and 5) evaluating the adequacy of the solution and, if necessary, restarting the cycle.

Following this analysis, the first- and third-year students' computing teachers, with the researchers' suggestions in mind, developed a task that could be used for assessment and that covered as much various facets as possible of the varying facets. (Some of the researchers are in fact computing teachers.) In designing the task, a requirement was that first-year students should encounter some difficulties in solving it but should be able to solve at least part of it; the task would be easier for third-year students. The goal was to measure the improvement in performance when the two academic levels were compared. The task had to include the problem-solving aspects given above. The Appendix gives the full task, with its problem statement and test questions. These questions map the scheme above: Questions 4, 8, 19, and 20 relate to aspect 2 (*creating with a strategy to solve the problem*); questions 2, 5, and 8 relate to aspect 4 (*applying knowledge needed for problem solving*); and questions 6, 13, and 14 relate to aspect 5 (*evaluating the adequacy of the solution*). It is important to emphasize that, although the problem statement has algorithmic aspects, it is not a programming problem. The only item that demands a programmed solution is question 2, whose solution is immediate, as shown as follows:

```

 $S := 0;$ 
For  $i = 1$  to  $n^2$  do  $S := S + 2^{i-1};$ 
Return  $S.$ 

```

TABLE II
STATISTICAL RELIABILITY

	Cronbach's alpha	N
Open questions	0.622	9
Closed questions	0.520	11
Total test	0.689	20

This question is worth a maximum of three points. The maximum score for the whole test is 38; hence, this question is only worth 7.9% of the entire test score.

The test that follows the problem statement (see the Appendix) has two types of questions, i.e., 9 open questions, requiring a constructive response, and 11 closed multiple-choice questions, requiring a selection of one of four alternative answers. Requiring constructive responses is a common practice in LO assessment since these are thought to better reflect the competences being measured, as the test taker has to respond without external hints [28]. The effort required to rate open questions, however, is higher, and inter-rater reliability is lower. The authors felt that a mix of the two question types could be advantageous and would also allow a comparison of how well they serve their purpose. A marking scheme was designed for the test. The open questions, 1–9, were worth three points each; the multiple-choice questions, 10–20, were worth one point each.

3) *Inter-Scorer Agreement*: First, the 71 student participants (see Table I) answered the task questions. Two Master's students were then selected to answer the task questions to correct them. They were then given ten of the 71 student responses to the test questions to grade, along with a rubric prepared by the instructors. Two instructors also graded these ten answer sets separately. The Master's students' and instructors' grade ratings were compared, the differences were analyzed, and the rubric was improved accordingly. Then, this new rubric was applied to a further set of 22 student answers, and the agreement of instructor and Master's student ratings was calculated using Cohen's kappa coefficient [29] to remove random agreement. This process continued until a satisfactory level of agreement was reached (Cohen's kappa was found to be 0.497). According to Landis and Koch [30], this is a moderate level of agreement that should be improved by revising the scoring criteria and the examples provided. All student answers were then graded to this final standard.

4) *Internal Consistencies of the Test*: Cronbach's alpha is a commonly used index for this feature, which reveals how the task focuses on what it intends to measure. Results show a value of 0.689. While a figure of 0.70 would be more desirable [31], this is close and could be considered acceptable. Table II shows Cronbach's alpha value for the answers to the open and closed questions and to the overall test.

5) *Content Validity of the Test*: The same task that was discussed in this paper was sent to two educational experts who were asked to identify which competence they tried to measure. Both agreed that the task measured problem-solving ability and met the goals of its designers. This suggests that the task has a clear apparent validity and that the content meets the purpose for which it was designed.

TABLE III
STATISTICAL PARAMETERS FOR THE TEST SUBSCALES
AND THE OVERALL TEST BY YEAR

	Year	N	Mean	SD	SE of Mean
Open questions	1	50	0.96	1.309	0.185
	3	21	2.90	2.448	0.534
Closed questions	1	50	3.80	1.807	0.256
	3	21	6.00	2.302	0.502
TOTAL test	1	50	4.76	2.344	0.331
	3	21	8.90	4.049	0.883

TABLE IV
(a) LEVENE'S TEST FOR EQUALITY OF VARIANCE;
(b) T-TEST FOR EQUAL MEANS

(a)

	F	Sig.	Conclusion
Sub-scale 1	15.566	0.000	No equality of variances
Sub-scale 2	2.137	0.148	Equality of variances
TOTAL	13.629	0.000	No equality of variances

(b)

	t	Sig. (bilateral)
Sub-scale 1	-3.441	0.002
Sub-scale 2	-4.309	0.000
TOTAL	-4.393	0.000

6) *Discriminant Validity*: For the purpose of the study, it was important that the test accurately distinguishes first- and third-year students since it may be assumed that the latter developed problem-solving competence during their studies. Table III summarizes the main statistical descriptors for the test scores for the first- and third-year students for the open and closed question subscales and for the overall test (TEST TOTAL). In all cases, mean values for third-year students are higher than the mean for first-year students, showing a significant development of problem-solving competence over time.

A T-test was performed to compare the mean values for the scores of subscale 1 (open questions) and subscale 2 (closed questions) and the overall test, by the two participant classes, to check for differences and, if any, to determine if they are statistically significant [see Table IV(a) and (b)].

Taking the significance level (Sig. bilateral) as a reference, in all cases (subscale 1—open questions; subscale 2—closed questions; and overall test), this level is less than 0.05 ($p < 0.0005$ for subscale 2 and the overall test and 0.002 for subscale 1); hence, empirical evidence was not found to indicate that means in grades are equal.

As shown in Table IV(b), the test yields results that significantly differ from the two academic levels. To analyze how big this difference is, the effect size for both subscales and for the overall test was calculated using Hedges' g value [32]. Since both samples (first- and third-year students) are independent and this is not an experimental study (i.e., with experimental and control groups), a pooled standard deviation was used [32]. Computations were carried out using online available software

TABLE V
EFFECT SIZE

Sub-scales	Effect size (g)
Open questions	1.116
Closed questions	1.108
TEST TOTAL	1.392

TABLE VI
EFFECT SIZE (G , D , AND POWER)

Sub-scales	Effect size (g)	Effect size (d)	Power
Open questions	1.116	1.132	0.9961525
Closed questions	1.108	1.124	0.9957896
TEST TOTAL	1.392	1.412	0.9999046

[33], [34]. Table V shows the g value for both subscales and the overall test.

From these results, it can be concluded that the effect size is very large (strong effect). It is also useful to compute the statistical power of the test that was used using the G*Power tool [35] (see Table VI).

It is shown that, in all cases, if H_0 is false, the probability of refusal is above 99.5%. The power is very high because the effect size is very high (>1 , in all cases), and the calculation includes both samples (50 and 21 students). For this calculation, " α err pro" value was set to the standard one, i.e., 0.05. If the effect size value were medium (0.5), with this sample, the power value of 0.6 would be obtained. This value is lower than 0.8, considering the minimum value for a correct interpretation from the results.

7) *Item Difficulty*: The test was difficult for all students. Given that the maximum possible score is 38 points (27 points for open questions and 11 points for closed questions), all three universities fall far below that score and even below its arithmetic mean (19). It may be concluded from the poor results that the test was difficult for these students or, alternatively, that the students have not developed this competence to the required level.

These results suggest how the test can be improved if it is to be adjusted to what the students can do; the more difficult items should be eliminated and possibly substituted by easier ones. However, they also suggest that the teachers should determine whether the students, particularly those in the third year, should have been able to respond to the questions they failed to answer and to what extent. That is, standards should be set. LO assessment is not based on normative data but is rather criterion-referenced; thus, these criteria must be carefully and clearly established.

III. DISCUSSION

Overall, the task developed is a valid tool for its purpose of measuring students' problem-solving competence. The presence of both open and closed (the most common) questions in the test, as recommended in the literature [28], should be emphasized. The open questions, however, complicated arriving at inter-scorer agreement. This illustrates one of the clear limitations of using open questions to assess competences, i.e., the difficulty of arriving at clear scoring criteria and the heavy human and time resources necessary to reach acceptable levels of precision. The training process, as well as the elaboration

of scoring criteria, is costly but unavoidable if objectivity in interpretation of results is to be guaranteed. Furthermore, a good level of inter-rater agreement cannot be taken for granted, but rather takes considerable effort. In general, the internal consistency of the test is close to the required value of 0.7 (0.689). Moreover, the test has a clear apparent validity and the content appears to fit the purpose it was designed for.

The discriminant validity of the test was validated, reflecting both that the assessment device is able to discriminate between different ability levels and that the problem-solving competence did develop over the three academic years. In addition, the difference in competence performance is remarkable (very large effect size) with a probability value higher than 99.5%.

Finally, results suggest that the complexity of the tests should be adjusted to match what students can do—a question if whether they should have been able to tackle the problems they were not able to solve. It should be underlined that, as stated in Section II-B2, the improved competence was in the ability to solve problems and not in the ability to program. Separate tests have been developed to assess programming skills.

It would be helpful to have a larger and more balanced sample of university students, with equal numbers enrolled in each academic year, to draw stronger conclusions about the differences between students at early and late stages of their university careers. Due to low participation, conclusions given in this paper must be treated with caution.

IV. CONCLUSION

At a time of high interest in competence-based higher education, reliable and valid devices to assess competences are needed. Developing procedures adequately able to assess transferable competences takes time and effort and is therefore costly [21], [25]. However, CBE cannot be said to have been put into practice if LOs are not assessed appropriately. This cannot be done, particularly for certification purposes, if it cannot be done objectively and in a valid way. Developing these measurement devices requires multidisciplinary knowledge and experience and thus calls for the collaboration of professionals from different disciplines.

The experience of this project raises several issues for discussion. The results are encouraging in two ways: The task developed seems to have good reliability and validity and is therefore suitable for the purpose for which it was created; more importantly, it offers many possibilities for expansion and improvement in the future. The more difficult questions have been identified, which indicates how the test can be improved in terms of its difficulty level and internal consistency. Whether it should be made easier or not depends very much on the expert judgment of teachers who must decide what students should be able to do at different academic levels.

The experience has been also extremely positive in other ways. It demonstrated the crucial value of assessment in educational practice. Teachers who collaborated with this project from the beginning, i.e., from the development of the assessment task, had a valuable experience that helped them understand the concept of competences and lay a solid foundation for CBE. Some students were observed to have difficulties

working with open tasks that they had to define for themselves, which broadly speaking is an important part of the problem-solving competence. It must be added that some teachers and academic authorities were reluctant to make time to include the task in their academic planning. In fact, even motivated teachers seemed to have difficulty incorporating this task as part of their work, which may underline the authors' earlier comments as to the complexity of CBE.

The main achievement of this project is the development of a procedure to measure some relevant LOs, specifically problem solving. However, as shown in this paper, this process takes much time and effort and is probably best approached as a multidisciplinary endeavor.

V. LIMITATIONS OF THE STUDY

A clear limitation of this study is that the student sample is very small, and this limits the generalization of its results. With a sample of this size, only classic psychometric analyses could be performed. A broader sample would allow analyses based on item response theory, which would provide more detailed data on each item, such as the information it adds to the test and the error it introduces. However, authors have been able to carry out these analyses with other student samples, part of a pilot project with a different but related task to that described here. The results show that the intermediate scoring levels (2 and 3) need to be carefully defined. In other words, the best and worst performance levels seem easy to identify, but the intermediate levels are fuzzier and should be defined more clearly.

APPENDIX TASK PROPOSED

PROBLEM STATEMENT

Legend has it that a king was so excited about chess that he offered the inventor of the game anything that he desired as a reward for his creation. After thinking about it for a while, the inventor claimed his reward: a grain of wheat for the first square of the chessboard, two grains for the second, four for the third, eight for the fourth, sixteen for the fifth, and so on. Imagine now that the chessboard has n boxes per side, instead of eight. Let $S(n)$ be the total wheat grains of reward for this chessboard.

QUESTIONS

1. Determine a general expression for $S(n)$.
2. Build an iterative program for computing $S(n)$.
3. What is recursion? Indicate those aspects that must appear in its definition. (Maximum 4 lines).
4. Indicate a general recursive strategy to find recursive solutions for a given problem P. (Maximum 10 lines).
5. Apply this strategy to find a recursive expression for $S(n)$. You must explain how each step of the strategy is applied for solving this problem, and the result of that application. (Maximum 20 lines).
6. It is said that a disadvantage of recursion is that its solutions are inefficient in terms of resource use. Give arguments for or against this statement. (Maximum 20 lines).

Given the next program code

```
dec A : num- > num;
  A(n) <= if n < 10
    then n mod 2
    else ((n mod 10) mod 2) + A (n div 10).
```

7. Indicate the set of base or output cases for this recursive scheme.
8. Indicate the recursive rule.
9. Obtain A (3475).

Mark the correct answer with a circle.

10. To differentiate between a recursive scheme and an iterative one, the key is to detect: a) the presence of self-references; b) the presence of iterative loops; c) the definition of the initial values; and d) the absence of a particular case.
11. In a recursive scheme, the set of base or output cases is: a) the set of initial values for the argument of the problem; b) a fixed-size set of argument values for which the solution to the problem is not recursively defined; c) a set of values for the argument of the problem for which the scheme provides the solution directly, without using simpler cases, i.e., they are used to terminate the recursive process; and d) a set of values for a variable that controls the termination of the process.
12. To correctly determine the set of base or output cases of a recursive scheme, it is necessary: a) to check that the first argument values of the problem are covered; b) to analyze the type of recursive rules of the scheme and, accordingly, to determine the size and elements of the base cases or output set; c) to set the simplest elements among the arguments of the problem; and d) to ensure that the set has the smallest possible size.
13. To verify that a given recursive scheme works correctly, we only have to: a) check that it has a set of base cases; b) verify that the recursive scheme reduces the size of the argument of the problem; c) check that, for certain argument values, the solution is obtained in a finite number of steps; and d) prove by induction that, in a finite number of applications of the recursive rules, the set of base or output cases is reached.
14. From the point of view of efficiency, if you had to choose between several recursive schemes that are applicable to the same problem, you would mainly consider: a) the number of steps or run required by the scheme to reach the set of base or output cases; b) the type of operations required by the recursive scheme; c) the number of subproblems in which the recursive scheme divides the original problem; and d) the size of the set of base or output cases.
15. The algorithmic complexity of recursive program


```
dec A : num- > num;
  A (n) <= if n < 10
    then n mod 2
    else ((n mod 10) mod 2) + A (n div 10)
```

 a) is logarithmic; b) is linear; c) is quadratic; or d) is exponential.
16. In the original chess problem, for the expression S(64), assuming that, for every grain put on chessboard, the computer has to print the corresponding ordinal number and considering that this impression takes 1 ms to complete: a) printing all grains takes a reasonable time (less than 1 h); b) the printing process will never end; c) the printing process takes longer than a year; and d) printing time is 64 ms.
17. When I am solving a problem, the first thing I do is: a) to explore different ways to solve it, until I find the best one; b) I do not usually waste time beating around before finding the solution; c) to try to understand it clearly; d) to critically evaluate the solution, to make sure that I have found the best one.
18. Teachers often return exercises with corrections that: a) I do not find very useful, and they are also depressing; b) I find very useful, as it is good to know where I have failed; c) are useful for the task to which they relate but are not useful for other deliverables; d) usually, I do not have time to stop and think about them.
19. Generally, when I am solving problems, trying to define a problem before working on it: a) it can only be done when I have gathered enough information; b) it has more academic interest than practical; c) it is a waste of time since it is often not possible before trying different solutions; d) it is essential to start working.
20. If I cannot understand a problem: a) I try to see it from different points of view; b) I leave it because if you do not find the solution at the beginning, you will not find it; c) I try to find and test different solutions; d) I try to read it again and again.

ACKNOWLEDGMENT

One of the authors of this work, Dr. C. Vizcarro, died in 2014. The authors want this work to serve as a tribute to her professional work and her personal qualities. Thanks, Carmen!

REFERENCES

- [1] L. Goff *et al.*, "Learning Outcomes Assessment A Practitioner's Handbook," Univ. Windsor, Windsor, ON, Canada, Centre for Teaching and Learning Reports, 2015. [Online]. Available: <http://scholar.uwindsor.ca/ctlreports/6>
- [2] ECTS Users Guide, Eur. Commission, Ottawa, ON, Canada, 2009. [Online]. Available: http://ec.europa.eu/education/lifelong-learning-policy/ects_en.htm
- [3] Ministerio de Educación y Ciencia (2007, Oct. 30). Real Decreto 1393/2007, de 29 de Octubre, por el que se Establece la Ordenación de las Enseñanzas Universitarias Oficiales, BOE Núm. 260, Oct. 30, 2007, pp. 44037–44048.
- [4] OECD, *Higher Education and Adult Learning, Testing Student and University Performance Globally: OECD's AHELO*, 2011. [Online]. Available: <http://www.oecd.org/education/highereducationandadultlearning/testingstudentanduniversityperformancegloballyoecdshelo.htm>
- [5] "Criteria for accrediting engineering programs," ABET Eng. Accreditation Commission, Baltimore, MD, USA, Oct. 2015. [Online]. Available: <http://www.abet.org/wp-content/uploads/2015/10/E001-16-17-EAC-Criteria-10-20-15.pdf>
- [6] S. A. Male, M. B. Bush, and E. S. Chapman, "An Australian study of generic competencies required by engineers," *Eur. J. Eng. Educ.*, vol. 36, no. 2, pp. 151–163, May 2011.

- [7] I. Markes, "A review of literature on employability skill needs in engineering," *Eur. J. Eng. Educ.*, vol. 31, no. 6, pp. 637–650, Dec. 2006.
- [8] B. S. Bloom, *Taxonomy of Educational Objectives, Handbook 1: The Cognitive Domain*. New York, NY, USA: David McKay, 1956.
- [9] B. S. Bloom, J. T. Hastings, and G. F. Madaus, *Handbook on Formative and Summative Evaluation of Student Learning*. New York, NY, USA: McGraw-Hill, 1971.
- [10] W. J. Popham, *Educational Tests and Measurements; Addresses, Essays, Lectures*. New York, NY, USA: Prentice-Hall, 1978.
- [11] Bowden & Marton, *Effective Learning and Teaching in Engineering*. New York, NY, USA: RoutledgeFalmer, 1998.
- [12] J. Stephenson, "The concept of capability and its importance in higher education," in *Capability and Quality in Higher Education*, J. Stephenson and M. Yorke, Eds. London, U.K.: Kogan, 1998, pp. 1–13.
- [13] S. Otter, "Learning outcomes in higher education," in *Outcomes, Learning and the Curriculum*, J. Burke, Ed. London, U.K.: Falmer, 1995.
- [14] D. R. Sadler, "Interpretations of criteria-based assessment and grading in higher education," *Assessment Eval. Higher Educ.*, vol. 30, no. 2, pp. 175–194, 2005.
- [15] D. Bushway and D. Everhart, "Investing in Quality Competency-Based Education (EDUCAUSE Review)|EDUCAUSE.edu. EDUCAUSE, Dec. 8, 2014. [Online]. Available: <http://www.educause.edu/ero/article/investing-quality-competency-based-education>
- [16] A. Edward and P. Knight, *Assessing Competences in Higher Education*. London, U.K.: Kogan, 1995.
- [17] R. S. Nickerson, D. N. Perkins, and E. E. Smith, *The Teaching of Thinking*. Hillsdale, NJ, USA: Lawrence Erlbaum Assoc., 1985.
- [18] G. Polya, *How to Solve It*. New York, NY, USA: Doubleday, 1957.
- [19] A. Newell and H. A. Simon, *Human Problem Solving*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1972.
- [20] A. H. Shoenfeld, "Teaching problem-solving skills," *Amer. Math. Monthly*, vol. 87, pp. 794–805, Oct. 1980.
- [21] J. E. Pérez Martínez, J. García, I. Muñoz, and A. Sierra, "Active learning and generic competences in an operating systems," *Int. J. Eng. Educ.*, vol. 26, pp. 1484–1492, Dec. 2010.
- [22] K. Tasa, S. Taggar, S. Seijts, and G. H. Seijts, "The development of collective efficacy in teams: A multi-level and longitudinal perspective," *J. Appl. Psychol.*, vol. 92, no. 1, pp. 17–27, Feb. 2007.
- [23] J. García and J. E. Pérez Martínez, "Measuring the influence of Cooperative Learning and Project Based Learning on problem solving skill," in *Proc. Res. Eng. Educ. Symp.*, Madrid, Spain, 2011, pp. 1–6.
- [24] P. P. Heppner, *The Problem Solving Inventory (PSI): Manual*, Univ. Missouri, Columbia, MO, USA, 1998.
- [25] C. Vizcarro Guarch *et al.*, "Assessment of learning outcomes in computing studies," in *Proc. IEEE Educon Conf.*, Berlin, Germany, 2013, pp. 486–491.
- [26] V. Wass, C. Van der Vleuten, J. Shatzer, and R. Jones, "Assessment of clinical competence," *Lancet*, vol. 357, pp. 945–949, Jan. 2001.
- [27] *Standards for educational and psychological testing*, American Educational Research Association, American Psychological Association & National Council on Measurement in Education. Washington, DC: American Psychological Association, 1999.
- [28] R. L. Johnson, J. A. Penny, and B. Gordon, "Assessing performance: Designing, scoring, and validating performance tasks," *J. Educ. Meas.*, vol. 46, pp. 474–477, Apr. 2009.
- [29] J. Cohen, "A coefficient of agreement for nominal tables," *Educ. Psychol. Meas.*, vol. 20, pp. 37–46, Jan. 1960.
- [30] J. Landis and G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–74, Jan. 1977.
- [31] M. R. Martínez Arias, M. J. Hernández Lloreda, and M. V. Hernández Lloreda, *Psicometría*. Madrid, Spain: Alianza Editorial, 2006.
- [32] Hedges and V. Larry, "Distribution theory for glass's estimator of effect size and related estimators," *J. Educ. Statis.*, vol. 26, no. 2, pp. 107–128, Summer 1981.
- [33] D. Walker, *David Walker's Effect Size Calculator*, 2007. [Online]. Available: <http://www.cedu.niu.edu/~walker/calculators/effect.asp>
- [34] P. D. Ellis, *Effect Size FAQs*, 2010. [Online]. Available: <http://www.effectsizefaq.com>
- [35] F. Faul, E. Erdfelder, A.-G. Lang, F. Faul, E. Erdfelder, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods*, vol. 39, pp. 175–191, 2007.

Jorge Pérez received the B.S. degree in computer science from the Universidad Carlos III de Madrid, Leganés, Spain, in 1999, and the Ph.D. degree from the Universidad Politécnica de Madrid, Madrid, Spain, in 2004.

From 1993 to 1997, he was a Secretary of the Faculty of Computer Sciences, Universidad Politécnica de Madrid, and from 2007 to 2014, he was the Headmaster of the Department of Applied Computer Science, Universidad Politécnica de Madrid. Since 1985, he has been also an Associate Professor with the same university. He is the coordinator of the DMAE-DIA educational innovation group and the Competencies and Active Learning in Engineering Education research group. He has taken part in the creation of several syllabuses for both grades and postgraduates and has supervised many research projects about educational innovation in engineering.

Dr. Pérez is a recipient of several awards from the Rector of the Universidad Politécnica de Madrid for educational innovation at the university.

Carmen Vizcarro was a Professor with the Universidad Autónoma de Madrid (UAM), Madrid, Spain, where she taught psychological assessment. She was the Director of the Centre for Learning and Instruction, UAM, in 1997–2001 and of the Unit for Educational Innovation and Quality, Universidad de Castilla-La Mancha, Albacete, Spain. She was the Founder of the Spanish network for university teaching (Red Estatal de Docencia Universitaria) and served as Coordinator and Editor of its review until 2005. A member of the Spanish Bologna Experts Team (2009–2013), she served as the Coordinator of several projects on assessment of learning outcomes funded by the Spanish Ministry of Education and participated in several European projects on quality in higher education [e.g., Network for Tertiary-Level Educators (NETTLE) and Réseau de Centres de Ressources de l'Enseignement Supérieur (RECRE@SUP)]. She participated in evaluations for the Institutional Evaluation Program of the European University Association.

Javier García received the B.S. degree in computer science and a Diploma of advanced studies from the Universidad Politécnica de Madrid, Madrid, Spain, in 2000 and 2007, respectively.

He is currently an Associate Professor of computer architecture and technology with the Universidad Politécnica de Madrid. He is a member of the DMAE-DIA educative innovation group that works on the development of new methodologies for learning/evaluation. The DMAE-DIA group received an educative innovation award from the Universidad Politécnica de Madrid in 2009. This group has developed several projects on active learning and general competencies in computer science. He is the author or coauthor of a significant number of papers related to teaching in engineering education, which have been published over the last years. His research focuses on real-time systems and remote sensing areas. Currently, he works on issues related to designing classroom activities and evaluating students' learning on problem-based learning.

Aurelio Bermúdez received the Ph.D. degree in computer science from in 2004.

He is currently an Associate Professor of computer architecture with the Department of Computer Systems, Universidad de Castilla-La Mancha, Albacete, Spain. Since 2011/2012, he has been the Teaching Coordinator of the computer engineering degree, being responsible for the correct development and assessment of the learning outcomes and skills included in the program. He has taken part in 25 research projects, coauthoring more than 50 publications in these areas. His research interests include dependability in high-speed interconnection networks, information processing in wireless sensor networks, integration of unmanned air vehicles in Internet of Things applications, and interconnection networks for avionic systems.

Ruth Cobos received the B.S. and Ph.D. degrees in computer science from the Universidad Autónoma de Madrid (UAM), Madrid, Spain, in 1997 and 2003, respectively.

She was a Postdoctoral Researcher with the Technical University of Munich, Munich, Germany, and with the University of Southampton, Southampton, U.K. She is currently an Associate Professor with the Department of Computer Science and Engineering, Higher Polytechnic School, UAM. She collaborated with and supervised more than ten projects on educational innovation. She has published articles related to software engineering and the teaching methods applied both to engineering and to other disciplines. Her main research interests include collaborative systems, distributed and collaborative knowledge management, social learning, blended learning, learning analytics, massive open online courses, and educational technologies.